The XML parser plugin
PRINTED MANUAL

© 1999-2010 AGG Software

XML parser plugin

© 1999-2010 AGG Software

All rights reserved. No parts of this work may be reproduced in any form or by any means - graphic, electronic, or mechanical, including photocopying, recording, taping, or information storage and retrieval systems - without the written permission of the publisher.

Products that are referred to in this document may be either trademarks and/or registered trademarks of the respective owners. The publisher and the author make no claim to these trademarks.

While every precaution has been taken in the preparation of this document, the publisher and the author assume no responsibility for errors or omissions, or for damages resulting from the use of information contained in this document or from the use of programs and source code that may accompany it. In no event shall the publisher and the author be liable for any loss of profit or any other commercial damage caused or alleged to have been caused directly or indirectly by this document.

Printed: 20.07.10

Publisher

AGG Software

Production

© 1999-2010 AGG Software
http://www.aggsoft.com

Table of Contents

Part 1 Introduction 1

Part 2 System requirements 1

Part 3 Installing XML parser 1

Part 4 Glossary 4

Part 5 User Manual 5

1 Analyzing and parsing data 5
2 XML nodes 6
3 Characters translation 8
4 Filter 9
5 Syntax of Regular Expressions 10
6 Data format 15

Part 6 Troubles? 16

1 Possible problems 16

© 1999-2010 AGG Software

1 Introduction

This plugins allows to extract a XML data from data flow and parse it to XML nodes and attributes. These values will be exported as parser items and you can use their values in data export plugins.

2 System requirements

The following requirements must be met for "XML parser" to be installed:

Operating system:

- Windows 95;
- Windows 95 OSR2;
- Windows 98;
- Windows Me;
- Windows NT4;
- Windows 2000;
- Windows XP (all editions) (both x86 and x64);
- Windows 2003 Server (both x86 and x64);
- Windows Vista (both x86 and x64).

Free disk space: Not less than 5 MB of free disk space is recommended.
Special access requirements: You should log on as a user with Administrator rights in order to install this module.

The main application (core) must be installed, for example, Advanced Serial Data Logger.

Notes for Microsoft Vista:

Since our software saves data to the registry and installs to the Program Files folder, the following requirements must be met:

1. You need Administrator rights to run and install our software
2. The shortcut icon of our software will be located on the desktop;
3. Windows Vista will ask for your confirmation to continue the installation.

NOTE: You can configure the user account only once in order not to see the above dialog box any more. Search Google for the solution of this problem.

3 Installing XML parser

1. Close the main application (for example, Advanced Serial Data Logger) if it is running;
2. Copy the program to your hard drive;
3. Run the module installation file with a double click on the file name in Windows Explorer;
4. Follow the instructions of the installation software. Usually, it is enough just to click the "Next" button several times;

© 1999-2010 AGG Software

5. Start the main application. The name of the module will appear on the "Modules" tab of the "Settings" window if it is successfully installed.

If the module is compatible with the program, its name and version will be displayed in the module list. You can see examples of installed modules on fig.1-2. Some types of modules require additional configuration. To do it, just select a module from the list and click the "Setup" button next to the list. The configuration of the module is described below.

You can see some types of modules on the "Log file" tab. To configure such a module, you should select it from the "File type" list and click the "Advanced" button.

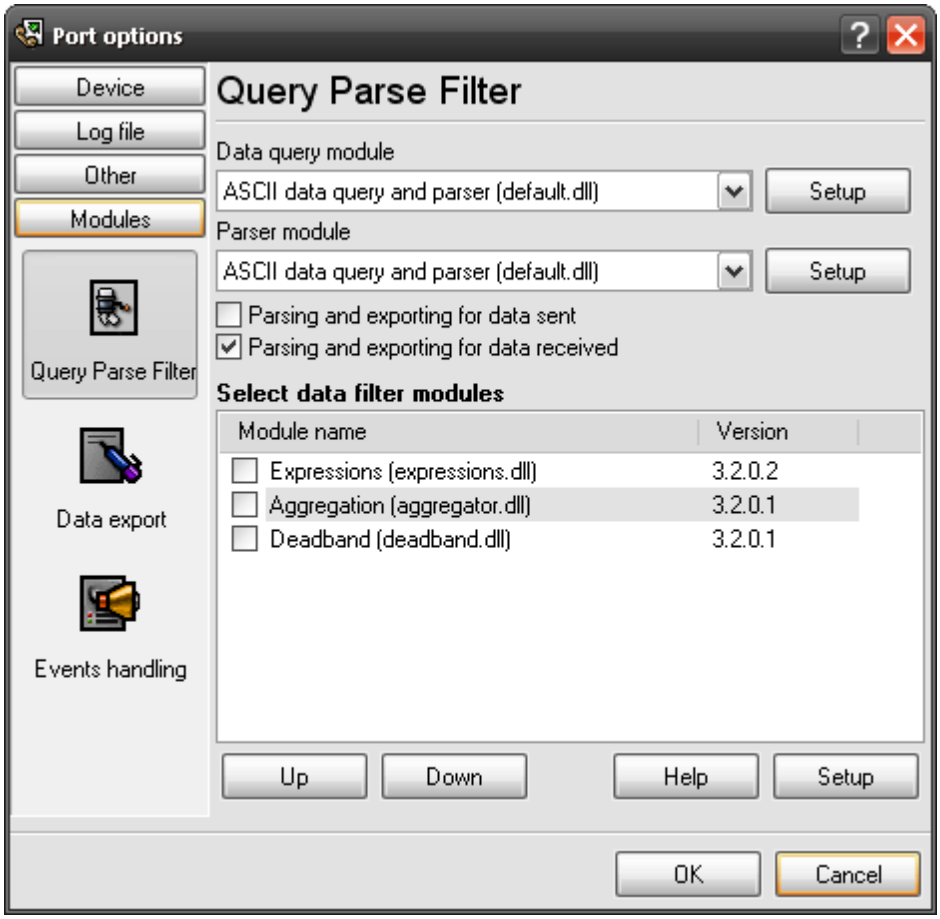

Fig.1. Examples of installed modules

© 1999-2010 AGG Software

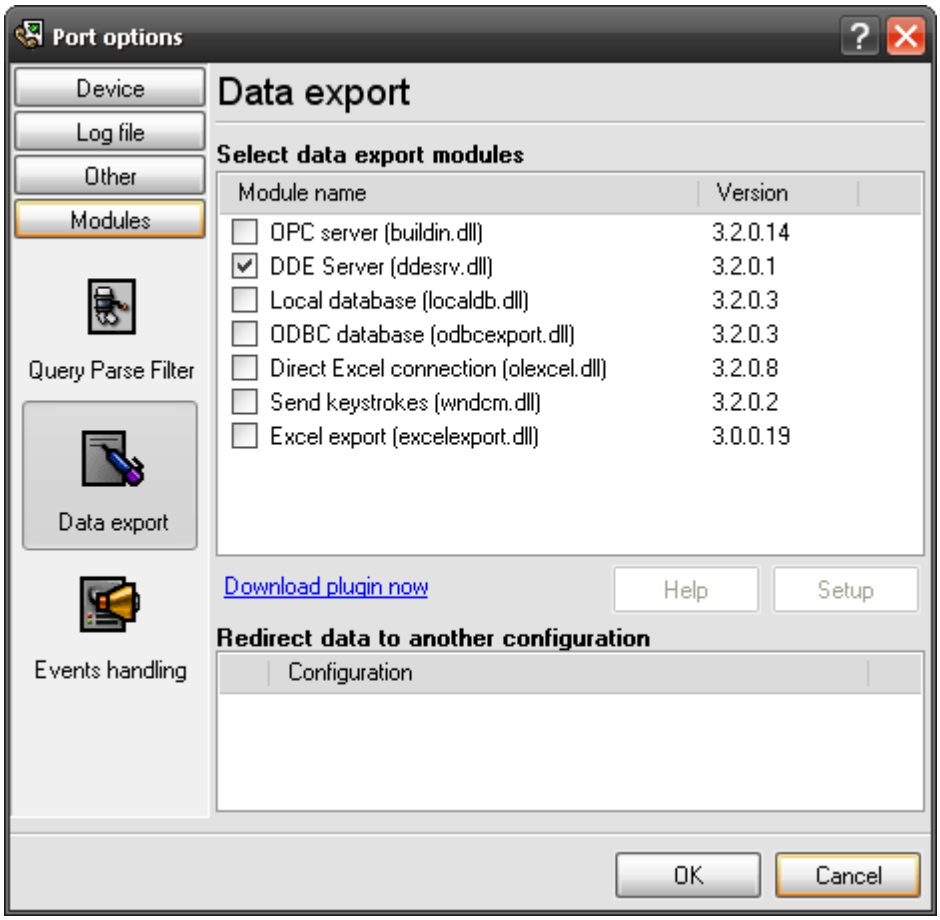

Fig.2. Examples of installed modules

© 1999-2010 AGG Software

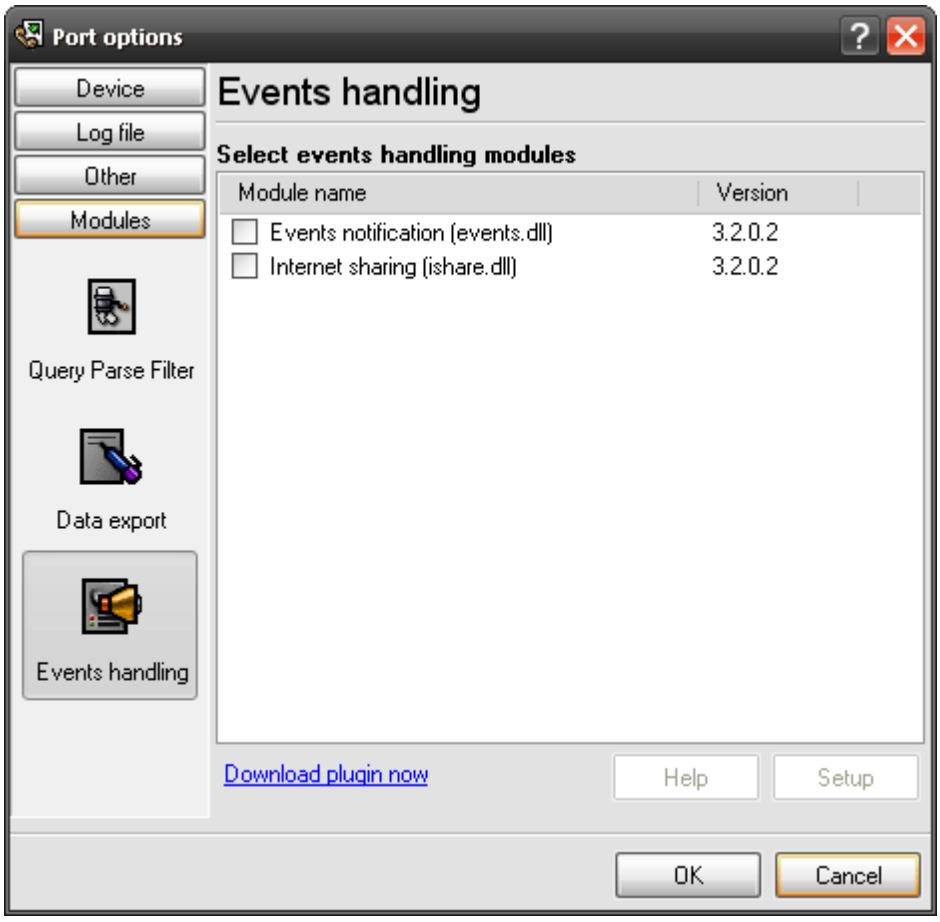

Fig.3. Examples of installed modules

4 Glossary

Plug-in - module

Main program – the program shell that uses this module. For example: Advanced Serial Data Logger

Parser – the module that processes the data flow singling out data packets from it and variables from data packets. These variables are used in data export modules after that.

Core - see "Main program".

© 1999-2010 AGG Software

5 User Manual

5.1 Analyzing and parsing data

To export the data received from the port, you should configure the parser. The XML parser allows you to single out XML nodes from the overall flow consisting of ASCII characters and parse these packets into XML nodes and attributes and their values that are exported into various destinations after that. You should define the rules the parser will use to process the incoming data flow on the "XML darata" tab (fig.1).

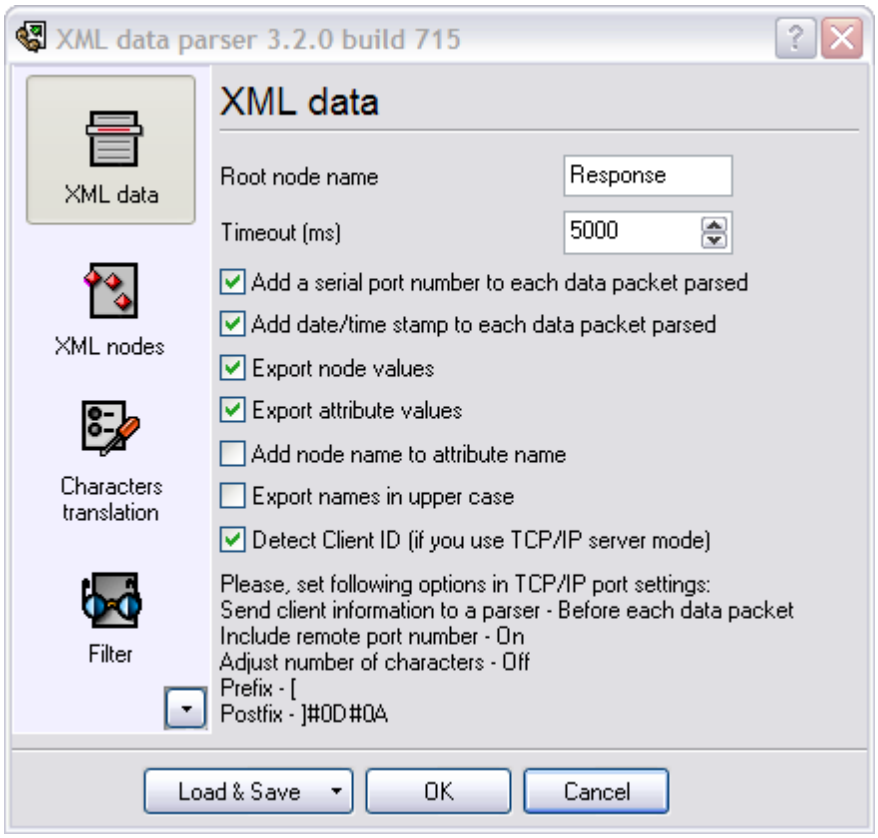
XML data parser 3.2.0 build 715
XML data
XML data
XML nodes
Characters translation
Filter
Root node name
Response
Timeout (ms)
5000
Add a serial port number to each data packet parsed
Add date/time stamp to each data packet parsed
Export node values
Export attribute values
Add node name to attribute name
Export names in upper case
Detect Client ID (if you use TCP/IP server mode)
Please, set following options in TCP/IP port settings:
Send client information to a parser - Before each data packet
Include remote port number - On
Adjust number of characters - Off
Prefix - [
Postfix -]#0D#0A
Load & Save
OK
Cancel

Fig.1. XML data.

Root node name – case insensitive name of the node that will contain all nodes for export. For example: Response;

The plugin will extract from the overall flow a XML data like:

```
<Response end="#0A" parse="bp" timeout="500" validatesize="0">
	<Item type="filter" size="-1" filtertype="1" exprtype="1" expr="----" />
	<Item type="filter" size="-1" filtertype="2" exprtype="2" expr="^\d{2}\/\d
{2}\/\d{2} " />
</Response>
```

All nodes inside <Response> and </Response>, and all attributes of <Response> will be exported.

© 1999-2010 AGG Software

If the XML node doesn't contain a termination tag (</Response>), then this XML data will not be extracted. For example, the following valid XML data will not be processed.

```
<Response end="#0A" parse="bp" timeout="500" validatesize="0">
....
</Response end="1">

<Response end="#0A" parse="bp" timeout="500" validatesize="0" />
```

Timeout – it may happen that the module will infinitely wait for the ending of some XML node. The timeout value is used to prevent it. It defines the maximum interval the module can wait for a XML node to end.

Add a time stamp to each data packet – the parser will add the new "data stamp" variable to each data packet that will be parsed into variables;

Add a port number to each data packet – the parser will add the new "port number" variable to each data packet that will be parsed into variables; It will allow you to identify data during export if you collect data from several devices at the same time.

Export node values – if this option is selected, then the plugin will export node values like: `<Item>Value</Item>`;

Export attribute values – if this option is selected, then the plugin will export values of all attributes of a node.

Example: `<Response end="#0A" parse="bp" timeout="500" validatesize="0">`;

attributes `end, parse, timeout, validatesize` will be exported.

Add node name to attribute name – if this option is selected, then a name of an node will be appended to an attribute name. For example above, attributes `end, parse, timeout, validatesize` will be exported with names `Response.end, Response.parse, Response.timeout, Response.validatesize`. It allows to identify same attributes of different XML nodes.

Export names in upper case – if this option is selected, then all names of attributes of nodes will be converted to upper case before export;

Detect Client ID – this option allows to use the parser with multiple TCP/IP clients, if you use our software in the server mode.

5.2 XML nodes

The "XML nodes" list (fig. 2) allows to define a data type for your XML nodes and attributes. Also you can specify to ignore some nodes and attributes. Click the "Add" button and and a new rule to the list.

© 1999-2010 AGG Software

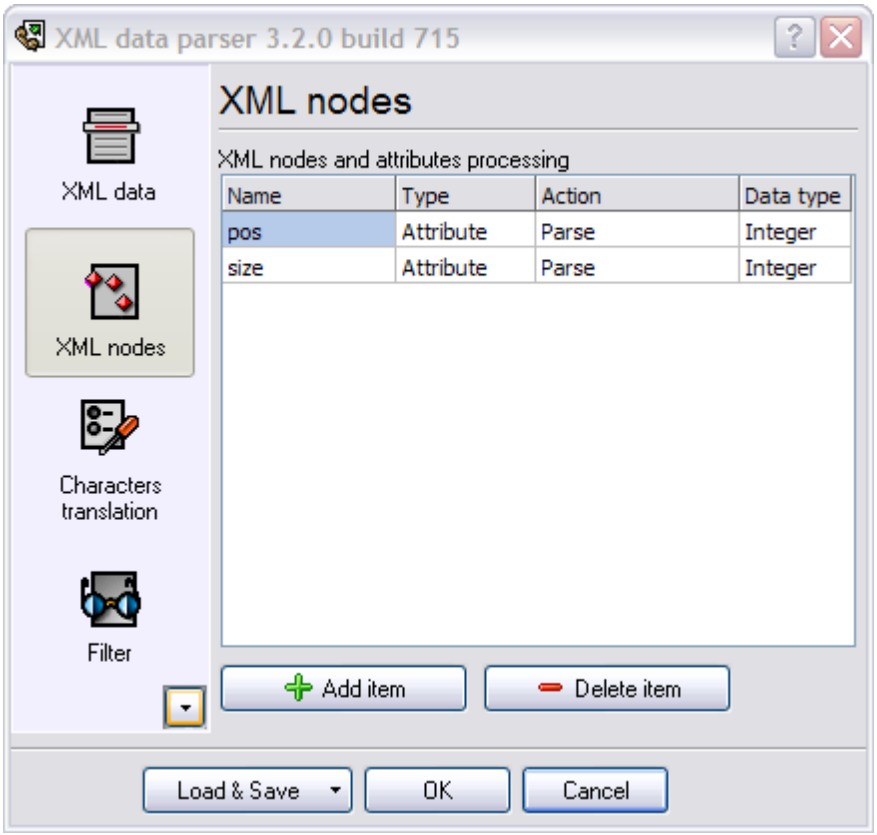

Fig.2. XML nodes.

.

Name - case insensitive name of an attribute or a node.

Type - the rule will be applied to All, Nodes or Attributes.

Action

- **Ignore** – the specified node or attribute will be ignored and will not be exported;
- **Parse** – the specified node or attribute will be parsed and exported.

Data type - the value will be converted from the String data type to the specified data type.

- String - String value - Character array with length from 1 to 65535 characters;
- Boolean - Logical value (True/False) - 0 or 1;
- Float - Real number - value range: -2.9 x 10^-39 .. 1.7 x 10^38
- Smallint - Signed small integer - value range: 32768..32767;
- Word - Word (unsigned small integer) - value range: 0..65535;
- Integer - Signed integer value: -2147483648..2147483647;
- Date - Date;
- Time - Time;
- DateTime - Date and time.

© 1999-2010 AGG Software

5.3 Characters translation

Character translation (fig.3) is used when you want to remove or replace some characters in a data packet. For example, remove nonprinting ASCII characters.

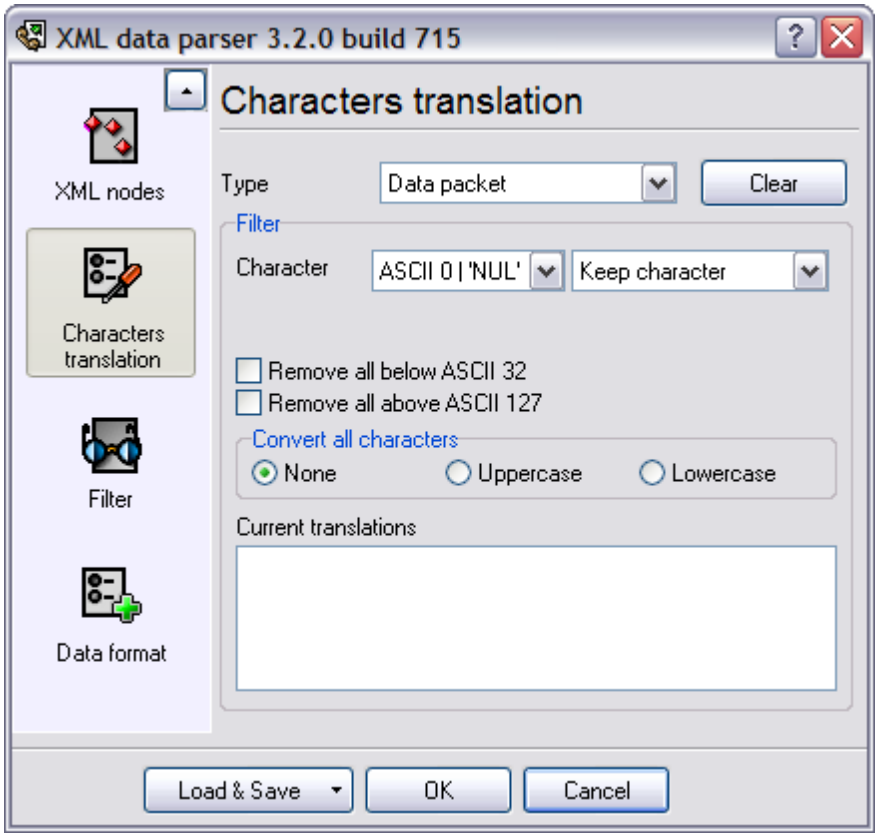

Fig.3 Character translation

© 1999-2010 AGG Software

5.4 Filter

The filter is used to ignore some data packets that you do not want to export with the help of other modules.

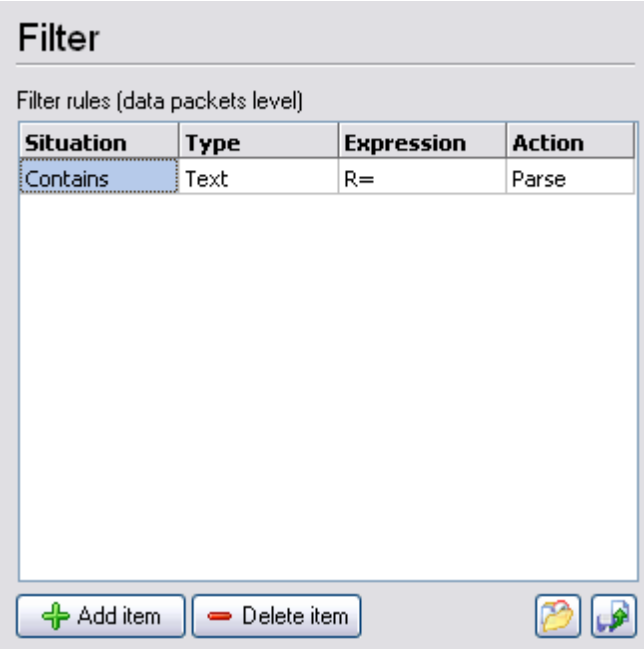

Fig.1 Filtering rules

You should define one or several filtering rules. If some condition is true, the program does what is specified in the "Action" field with the matching packet.

Action types

- **Ignore** – the current data packet will be ignored and will not be exported;
- **Parse** – the current data packet will be parsed and exported.

There are several condition types that are specified in the "Status" field.

Rule status types

- **Disabled** – this rule is disabled and not used for filtering purposes;
- **Contains** – this rules checks whether the string/expression from the "Expression" field is present in the data packet;
- **Does not contain** – this rules checks whether the string/expression from the "Expression" field is absent in the data packet.

Expressions types – Expressions in the "Expression" field can be of 2 types:

- **Text** – the module will search the data packet for a string specified in the "Expression" field. The search is case-sensitive.
- **Regular expression** – the module will use the regular expression specified in the "Expression" field in its search. The search is case-sensitive.

© 1999-2010 AGG Software

5.5 Syntax of Regular Expressions

Introduction

Regular Expressions are a widely-used method of specifying patterns of text to search for. Special **metacharacters** allow You to specify, for instance, that a particular string You are looking for occurs at the beginning or end of a line, or contains **n** recurrences of a certain character.

Regular expressions look ugly for novices, but really they are very simple, handy and powerful tool.

Let's start our learning trip!

Simple matches

Any single character matches itself, unless it is a **metacharacter** with a special meaning described below.

A series of characters matches that series of characters in the target string, so the pattern "bluh" would match "bluh" in the target string. Quite simple, eh ?

You can cause characters that normally function as **metacharacters** or **escape sequences** to be interpreted literally by 'escaping' them by preceding them with a backslash "\", for instance: metacharacter "^" match beginning of string, but "\^" match character "^", "\\" match "\" and so on.

Examples:

f oobar	*matchs string 'foobar'*
\ ^FooBar Pt r	*matchs '^FooBarPtr'*

Escape sequences

Characters may be specified using a **escape sequences** syntax much like that used in C and Perl: "\n" matches a newline, "\t" a tab, etc. More generally, \xnn, where nn is a string of hexadecimal digits, matches the character whose ASCII value is nn. If You need wide (Unicode) character code, You can use '\x{nnnn}', where 'nnnn' - one or more hexadecimal digits.

\ xnn	*char with hex code nn*
\ x{ nnnn}	*char with hex code nnnn (one byte for plain text and two bytes for Unicode)*
\ t	*tab (HT/TAB), same as \x09*
\ n	*newline (NL), same as \x0a*
\ r	*car.return (CR), same as \x0d*
\ f	*form feed (FF), same as \x0c*
\ a	*alarm (bell) (BEL), same as \x07*
\ e	*escape (ESC), same as \x1b*

Examples:

f oo\ x20bar	*matchs 'foo bar' (note space in the middle)*
\ t f oobar	*matchs 'foobar' predefined by tab*

© 1999-2010 AGG Software

Character classes

You can specify a **character class**, by enclosing a list of characters in [], which will match any **one** character from the list.

If the first character after the "[" is "^", the class matches any character **not** in the list.

Examples:

f oob[aei ou] r	*finds strings 'foobar', 'foober' etc. but not 'foobbr', 'foobcr' etc.*
f oob[^aei ou] r	*find strings 'foobbr', 'foobcr' etc. but not 'foobar', 'foober' etc.*

Within a list, the "-" character is used to specify a **range**, so that a-z represents all characters between "a" and "z", inclusive.

If You want "-" itself to be a member of a class, put it at the start or end of the list, or escape it with a backslash. If You want ']' you may place it at the start of list or escape it with a backslash.

Examples:

[- az]	*matchs 'a', 'z' and '-'*
[az-]	*matchs 'a', 'z' and '-'*
[a\ - z]	*matchs 'a', 'z' and '-'*
[a- z]	*matchs all twenty six small characters from 'a' to 'z'*
[\ n- \ x0D]	*matchs any of #10,#11,#12,#13.*
[\ d- t]	*matchs any digit, '-' or 't'.*
[ ] - a]	*matchs any char from ']'..'a'.*

Metacharacters

Metacharacters are special characters which are the essence of Regular Expressions. There are different types of metacharacters, described below.

Metacharacters - line separators

^	*start of line*
$	*end of line*
\ A	*start of text*
\ Z	*end of text*
.	*any character in line*

Examples:

^f oobar	*matchs string 'foobar' only if it's at the beginning of line*
f oobar $	*matchs string 'foobar' only if it's at the end of line*
^f oobar $	*matchs string 'foobar' only if it's the only string in line*
f oob. r	*matchs strings like 'foobar', 'foobbr', 'foob1r' and so on*

The "^" metacharacter by default is only guaranteed to match at the beginning of the input string/ text, the "$" metacharacter only at the end. Embedded line separators will not be matched by "^" or "$".
You may, however, wish to treat a string as a multi-line buffer, such that the "^" will match after any

© 1999-2010 AGG Software

line separator within the string, and "$" will match before any line separator.

The "." metacharacter by default matches any character.

Note that "^.*$" (an empty line pattern) doesnot match the empty string within the sequence \x0D\x0A, but matchs the empty string within the sequence \x0A\x0D.

Metacharacters - predefined classes

\ w	*an alphanumeric character (including "_")*
\ W	*a nonalphanumeric*
\ d	*a numeric character*
\ D	*a non-numeric*
\ s	*any space (same as [\t\n\r\f])*
\ S	*a non space*

You may use \w, \d and \s within custom **character classes**.

Examples:

f oob\ dr *matchs strings like 'foob1r', "foob6r' and so on but not 'foobar', 'foobbr' and so on*

f oob[\ w\ s] r *matchs strings like 'foobar', 'foob r', 'foobbr' and so on but not 'foob1r', 'foob=r' and so on*

Metacharacters - iterators

Any item of a regular expression may be followed by another type of metacharacters - **iterators**. Using this metacharacters You can specify number of occurences of previous character, **metacharacter** or **subexpression**.

*	*zero or more ("greedy"), similar to {0,}*
+	*one or more ("greedy"), similar to {1,}*
?	*zero or one ("greedy"), similar to {0,1}*
{ n}	*exactly n times ("greedy")*
{ n, }	*at least n times ("greedy")*
{ n, m}	*at least n but not more than m times ("greedy")*
* ?	*zero or more ("non-greedy"), similar to {0,}?*
+?	*one or more ("non-greedy"), similar to {1,}?*
??	*zero or one ("non-greedy"), similar to {0,1}?*
{ n} ?	*exactly n times ("non-greedy")*
{ n, } ?	*at least n times ("non-greedy")*
{ n, m} ?	*at least n but not more than m times ("non-greedy")*

So, digits in curly brackets of the form {n,m}, specify the minimum number of times to match the item n and the maximum m. The form {n} is equivalent to {n,n} and matches exactly n times. The form {n,} matches n or more times. There is no limit to the size of n or m, but large numbers will chew up more memory and slow down r.e. execution.

If a curly bracket occurs in any other context, it is treated as a regular character.

© 1999-2010 AGG Software

Examples:

f oob. * r	*matchs strings like 'foobar', 'foobalkjdflkj9r' and 'foobr'*
f oob. +r	*matchs strings like 'foobar', 'foobalkjdflkj9r' but not 'foobr'*
f oob. ?r	*matchs strings like 'foobar', 'foobbr' and 'foobr' but not 'foobalkj9r'*
f ooba{ 2} r	*matchs the string 'foobaar'*
f ooba{ 2, } r	*matchs strings like 'foobaar', 'foobaaar', 'foobaaaar' etc.*
f ooba{ 2, 3} r	*matchs strings like 'foobaar', or 'foobaaar' but not 'foobaaaar'*

A little explanation about "greediness". "Greedy" takes as many as possible, "non-greedy" takes as few as possible. For example, 'b+' and 'b*' applied to string 'abbbbc' return 'bbbb', 'b+?' returns 'b', 'b*?' returns empty string, 'b{2,3}?' returns 'bb', 'b{2,3}' returns 'bbb'.

Metacharacters - alternatives

You can specify a series of **alternatives** for a pattern using "|" to separate them, so that fee|fie|foe will match any of "fee", "fie", or "foe" in the target string (as would f(e|i|o)e). The first alternative includes everything from the last pattern delimiter ("(", "[", or the beginning of the pattern) up to the first "|", and the last alternative contains everything from the last "|" to the next pattern delimiter. For this reason, it's common practice to include alternatives in parentheses, to minimize confusion about where they start and end.
Alternatives are tried from left to right, so the first alternative found for which the entire expression matches, is the one that is chosen. This means that alternatives are not necessarily greedy. For example: when matching foo|foot against "barefoot", only the "foo" part will match, as that is the first alternative tried, and it successfully matches the target string. (This might not seem important, but it is important when you are capturing matched text using parentheses.)
Also remember that "|" is interpreted as a literal within square brackets, so if You write [fee|fie|foe] You're really only matching [feio|].

Examples:

f oo(bar | f oo)	*matchs strings 'foobar' or 'foofoo'.*

Metacharacters - subexpressions

The bracketing construct (...) may also be used for define r.e. subexpressions.

Subexpressions are numbered based on the left to right order of their opening parenthesis.
First subexpression has number '1'

Examples:

(f oobar) { 8, 10}	*matchs strings which contain 8, 9 or 10 instances of the 'foobar'*
f oob([0- 9] | a+) r	*matchs 'foob0r', 'foob1r' , 'foobar', 'foobaar', 'foobaar' etc.*

Metacharacters - backreferences

Metacharacters \1 through \9 are interpreted as backreferences. \<n> matches previously matched **subexpression** #<n>.

Examples:

(.) \ 1+	*matchs 'aaaa' and 'cc'.*

© 1999-2010 AGG Software

(. +) \ 1+ also match 'abab' and '123123'
([' "] ?) (\ d+) \ 1 matchs "'13" (in double quotes), or '4' (in single quotes) or 77 (without quotes) etc

Modifiers

Modifiers are for changing behaviour of parser.

There are many ways to set up modifiers.
Any of these modifiers may be embedded within the regular expression itself using the (?...) construct.

i
Do case-insensitive pattern matching (using installed in you system locale settings).
m
Treat string as multiple lines. That is, change "^" and "$" from matching at only the very start or end of the string to the start or end of any line anywhere within the string.
s
Treat string as single line. That is, change "." to match any character whatsoever, even a line separators, which it normally would not match.
g
Non standard modifier. Switching it Off You'll switch all following operators into non-greedy mode (by default this modifier is On). So, if modifier /g is Off then '+' works as '+?', '*' as '*?' and so on
x
Extend your pattern's legibility by permitting whitespace and comments (see explanation below)
.

The modifier /x itself needs a little more explanation. It tells the parser to ignore whitespace that is neither backslashed nor within a character class. You can use this to break up your regular expression into (slightly) more readable parts. The # character is also treated as a metacharacter introducing a comment, for example:

```
(
(abc)  # comment 1
|      # You can use spaces to format r.e. - parser ignores it
(efg)  # comment 2
)
```

This also means that if you want real whitespace or # characters in the pattern (outside a character class, where they are unaffected by /x), that you'll either have to escape them or encode them using octal or hex escapes. Taken together, these features go a long way towards making regular expressions text more readable.

How to change modifiers

(?imsxr-imsxr)
You may use it into r.e. for modifying modifiers by the fly. If this construction inlined into subexpression, then it effects only into this subexpression

© 1999-2010 AGG Software

Examples:

(?i) New- Yor k	*matchs 'New-york' and 'New-York'*
(?i) New- (?- i) Yor k	*matchs 'New-York' but not 'New-york'*
(?i) (New-) ?Yor k	*matchs 'New-york' and 'new-york'*
((?i) New-) ?Yor k	*matchs 'New-York', but not 'new-york*

(?#text)
A comment, the text is ignored. Note that parser closes the comment as soon as it sees a ")", so there is no way to put a literal ")" in the comment.

5.6 Data format

You can use the following tab to specify the data format for some data types (fig.5).

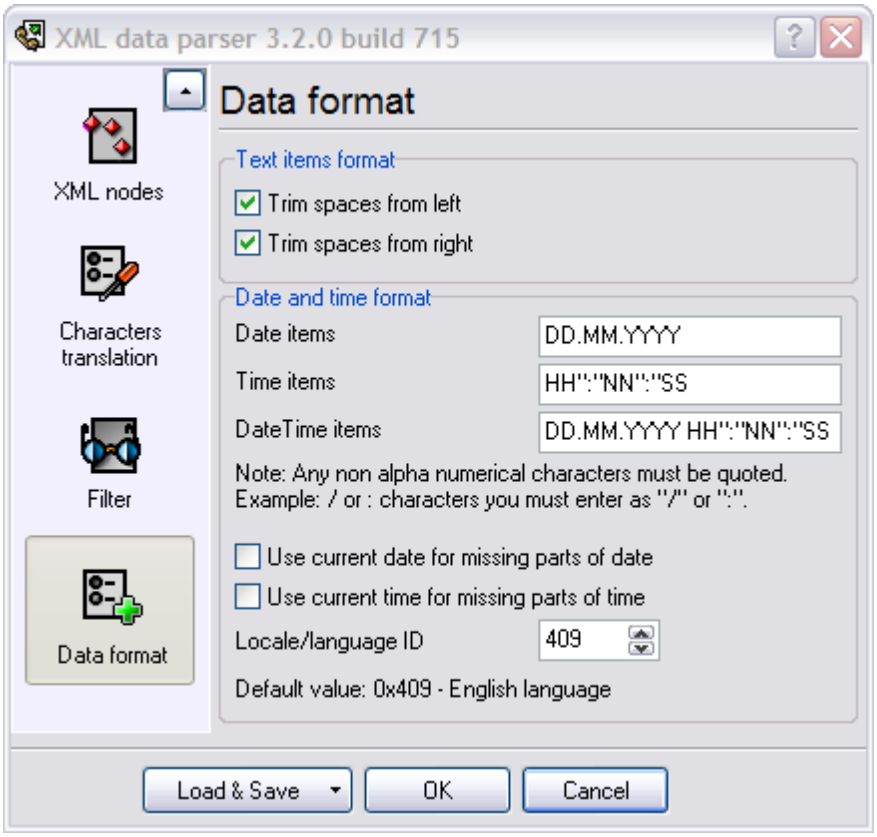

Fig.5. Data format

Text data format – this group of options allows you to remove leading and/or trailing spaces in variables of the String type.

Date and time format – it often happens that the date or time format the external device sends does not coincide with the format in which this data is stored in the database. To convert the date

© 1999-2010 AGG Software

into the necessary format, you should specify the format the date and time is received in.

The same format is used to specify the date and time as the one that is used in the name of the log file in the main program (for example, Advanced Serial Data Logger). So read about it in the help file of the main program. Note: the "/" and/or ":" characters are enclosed in quotation marks in the template (see fig.4).

If some part is missing in the received date and time (for example, year), you can take the missing part from the current date and time:

Use the current date for missing parts;
Use the current time for missing parts;

Sometimes devices send dates with month names (for example, "Jan, 10 2005") and this name can be in a language (in this example, it is in English) different from the language of your operating system. The language ID the date is sent in is specified in the "**Date language ID**" field in this case.

6 Troubles?

6.1 Possible problems

No data for publication/exporting – no data is passed for exporting. Solution: configure the parser, make sure that one or more variables are declared in the parser.

Error on binding variable with name %s [%s] – the error usually occurs if data does not correspond to the specified format. For example, the date and time format does not correspond to the data.

Unable to disconnect from the database [%s] and **Unable to connect to a database [%s]** – it is impossible to connect/disconnect to/form the database. You should check the parameters of the database connection. The analysis of the additional information will help you locate the error.

Database access error [%s]. Stop operations with the database? – the message appears if an error occurs during an attempt to execute an SQL query if the second variant of reacting to errors is selected. The message implies a "Yes" or "No" answer. The analysis of the additional information will help you locate the error.

Unable to verify your SQL script [%s] – the message appears when an attempt to analyze your SQL query fails. Check if the syntax of your SQL query is correct.

Tested successfully – the message appears if your database connection is successfully tested. It requires no additional actions.

Database isn't used – the message appears if the module is temporarily disabled (the "Temporarily disabled" check box is selected) or the database name field is empty. Check the connection parameters.

Database isn't selected - the message appears if the database type is not selected. Check the connection parameters.

© 1999-2010 AGG Software

Database: %s – %s contains the database name. The message appears if the database connection is successful. Usually, you see it when you call the module for the first time. It requires no additional actions.

Invalid data block length (columns=%d,length=%d) – an internal application error. It means that the data sent by the parser is in an invalid format. Perhaps, you are using the module incompatible with the version of the Advanced Serial Data Logger kernel. Update the versions of both the kernel and the module.

The time of connection is not due yet (%d,%d) – the message appears during an attempt to connect to the database after the connection to it has been lost and the "Reconnect after" option is enabled. No additional actions are required.

Invalid procedure call. Bad arguments –an attempt to call the module using invalid parameters. Perhaps, you are using the module incompatible with the version of the Advanced Serial Data Logger kernel. Update the versions of both the kernel and the module.

Writing to the database is complete - the message appears if your queue of SQL queries is successfully executed. It requires no additional actions.

Writing to the database is complete with errors – the message appears if the executing your queue of SQL queries was interrupted by an error. It requires no additional actions.

Your SQL is empty. Please, specify some SQL text first – the message appears if you do not enter the text for your SQL query. Check if the options on the "SQL queue" tab are configured correctly.

Invalid temporary path – the path to the temporary file specified by you does not exist. Enter a new path in the "Temporary folder" field on the "Errors handling" tab.

%s, %d – will be replaced by additional information.

© 1999-2010 AGG Software